



# Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru

Parawystia Prabasini Haryoto\*, Harly Okprana, Ilham Syahputra Saragih

Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: parawystia11@gmail.com

**Abstrak**—Melakukan klasifikasi kelulusan dari calon mahasiswa baru pada sebuah perguruan tinggi sebaiknya tidak hanya berdasarkan kriteria nilai ujian tertulis saja, tetapi juga test wawancara, dan hal lainnya yang dapat dijadikan parameter penilaian. Pada penelitian ini dilakukan penambahan parameter untuk mengklasifikasikan calon mahasiswa di STIKOM Tunas Bangsa Pematangsiantar dengan metode *deep learning* menggunakan algoritma C4.5. Dengan metode klasifikasi ini, diharapkan dapat membantu pihak akademik dalam menentukan kriteria calon mahasiswa baru yang aktif dan berprestasi. Dari hasil percobaan menggunakan software *RapidMiner* terhadap data mahasiswa yang telah mendaftar dari tahun 2016 sampai dengan tahun 2020, diperoleh klasifikasi mahasiswa yang aktif di tentukan oleh nilai wawancara yang menjadi node pertama, diikuti dengan nilai test potensi akademik. Sementara itu, pada node berikutnya ditemukan bahwa yang dapat mempengaruhi prestasi mahasiswa adalah asal sekolah dan lamanya masuk kuliah. Mahasiswa yang melanjutkan studi => 4 tahun cenderung memiliki nilai dan prestasi dibawah mahasiswa yang melanjutkan studi ke perguruan tinggi 0 sampai 4 tahun. Dasri hasil penelitian, diperoleh nilai akurasi sebesar 81.32%.

**Kata Kunci:** Data Mining; Klasifikasi Data; Algoritma C4.5; Pohon Keputusan; Penerimaan Mahasiswa Baru

**Abstract**—Conducting a graduation classification of prospective freshmen at a college should be best not only by the written exam value criteria but also the interview test, and other things that can serve as assessment parameters. In this study an increased parameters for classifying potential students in the scarlet hyperlinate's post-collegiate bud with deep learning methods using a c4 algoritmc.5. With this method of classification, it is hoped to help academicians determine the criteria of active and exemplary freshman candidates. From experiment with the rapidminer's software on the data of students who have registered from 2016 to 2020, obtained an active student classification defined by the value of the interview being the first node, coupled with the value of an academic potential test. In the meantime, it is found that what can affect a student's performance is the school's origin and duration of college. Students who continue studying => 4 years tend to have grades and achievements under those who continue to study at three to four years. Based on research, score of accuracy at 81.32%.

**Keywords:** Data Mining; Data Classification; C4.5 Algorithm; Decision Tree; Admission of New Students

## 1. PENDAHULUAN

Sekolah Tinggi Ilmu Komputer (STIKOM) Tunas Bangsa merupakan salah satu perguruan tinggi swasta di kota pematangsiantar yang memiliki banyak peminatnya. Pada saat ini STIKOM Tunas Bangsa memiliki dua program studi yaitu Sistem Informasi (SI) dan Teknik Informatika (TI). Setiap tahunnya STIKOM Tunas Bangsa membuka penerimaan calon mahasiswa baru, namun dalam penerimaan calon mahasiswa baru memiliki batas kuota. Banyaknya jumlah calon mahasiswa baru yang mendaftar membuat kampus sampai melebihi batas kuota penerimaan yang telah di tentukan. Untuk dapat mengetahui berapa jumlah mahasiswa yang masuk dan yang akan lulus, maka perlu di lakukan serangkaian cara untuk dapat mengetahuinya. Dengan menggunakan algoritma C4.5 maka akan dilakukan pengambilan informasi yang hasil penelitiannya dapat dijadikan contoh dalam masalah pengambilan keputusan yang melibatkan data-data yang pasti. Selama ini informasi yang diberikan dalam menentukan penerimaan calon mahasiswa baruhnya bisa dilihat berdasarkan Nilai Tes Seleksi Penerimaan Mahasiswa Baru, yang menunjukkan kualitas calon mahasiswa baru secara umum tanpa mengetahui keahlian yang benar-benar dimiliki oleh calon mahasiswa tersebut. *Data mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Untuk menerapkan konsep *data mining* banyak cara yang dapat digunakan diantaranya adalah menggunakan konsep beberapa algoritma yang ada pada *data mining*. Algoritma C4.5 adalah salah satu algoritma yang digunakan untuk melakukan klasifikasi data dengan membentuk pohon keputusan. Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*. Pemanfaatan atau penerapan *data mining* tersebut menjadi salah satu elemen yang terpenting untuk dicermati perkembangannya. Tujuan dari semua itu hanyalah untuk memudahkan pemakai dalam melaksanakan pekerjaan dan mengkomidasi setiap bentuk kebutuhannya.

Penelitian sebelumnya, Dalam penelitian ini yang berjudul “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu), Dari hasil penelitian terbukti bahwa algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analis mahasiswa. Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa tingkat ketepatan waktu mahasiswa menyelesaikan masa studinya [1]. Penelitian sebelumnya, Aplikasi data mining dengan metode klasifikasi algoritma C4.5 ini dapat memproses data mahasiswa menjadi sebuah aturan-aturan yang berguna sebagai masukan penentu keputusan walaupun jumlah data dapat mempengaruhi persentase keakuratannya. Berdasarkan hasil evaluasi dengan membandingkan keputusan asli dengan keputusan hasil mining C4.5, maka dapat dihitung persentase keakuratan kesesuaian sehingga memperoleh nilai sebesar 94% [2]. Dalam implementasi data mining dengan algoritma *decision tree* C4.5 untuk prediksi kelulusan mahasiswa menghasilkan model dan rule. Dan dari hasil model kelulusan mahasiswa tersebut dilakukan evaluasi algoritma sehingga menghasilkan nilai akurasi 65,98% dan termasuk klasifikasi data Baik [3].

Metode C4.5 yang digunakan dalam klasifikasi mahasiswa yang berpotensi drop out menghasilkan tujuh belas (17) rule. 17 rule tersebut dapat dijadikan sebagai pola dalam menentukan mahasiswa yang berpotensi mengalami drop out [4]. Penggunaan algoritma C4.5 dapat membantu pihak resto dalam meningkatkan pelayanan sesuai dengan hasil kuisioner yang telah diisi oleh para pelanggan. Berdasarkan hasil perhitungan, didapatkan nilai gain tertinggi adalah variabel tangibles dengan nilai 0,378118 dan untuk indikatornya pelanggan menyatakan puas terhadap Fasilitas, Peralatan, dan Penampilan Karyawan. Sedangkan pelanggan tidak puas terhadap satu indikator yaitu Komunikasi. Memudahkan pihak resto dalam mengumpulkan data dari pelanggan untuk meningkatkan pelayanan. Bisa untuk memberikan masukan kepada Sunrise Resto & Cafe dalam meningkatkan pelayanan kepada pelanggan[5].

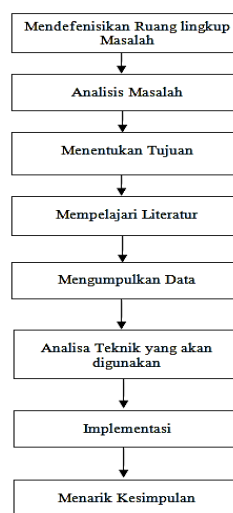
Berdasarkan pada rumusan masalah yang diteliti, proses Data Mining dengan menggunakan metoda Algoritma C4.5 dalam mengidentifikasi faktor-faktor kecelakaan kerja konstruksi maka dapat diambil kesimpulan bahwa metode Algoritma C4.5 atau pohon keputusan lebih efektif dan fleksibel jika digunakan pada proses pengklasifikasian. Sistem yang diimplementasikan menggunakan salah satu metode Data Mining algoritma C4.5 dan telah diuji dengan software Data Mining Weka Gui Chooser dalam mengidentifikasi faktor-faktor kecelakaan kerja konstruksi memiliki hasil pengklasifikasian sama[6]. Algoritma C4.5 dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan (decision tree) maupun dalam bentuk aturan atau rule If – Then sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan [7]. Sistem yang dibangun dapat membantu dalam pengklasifikasian tingkat keganasan hama tanaman padi yang umumnya masih dilakukan secara manual. Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi eksekutif dan sistem digunakan untuk menggambarkan proses yang terkait dengan pengklasifikasian keganasan hama tanaman padi. Sistem pengklasifikasian tingkat keganasan hama tanaman padi menggunakan algoritma C4.5 dapat digunakan dalam pengambilan keputusan untuk mencari alternatif yang baik[8]. Klasifikasi proses seleksi calon peserta lomba siswa SMP dapat mengklasifikasikan siswa dalam tahapan lolos atau tidaknya dalam seleksi. Dari 33 data siswa yang digunakan menunjukkan tingkat akurasi dengan algoritma C4.5 sebesar 81,81% [9].

Dari hasil penerapan algoritma C4.5 dalam prediksi kelulusan mahasiswa Prodi Informatika dapat disimpulkan bahwa atribut yang paling dominan dalam kelulusan mahasiswa adalah IPK, kedua adalah TOEFL, ketiga adalah asal daerah, dan yang terakhir adalah jenis kelamin. Pemotongan pohon keputusan pada algoritma C4.5 dapat meningkatkan akurasi. Pohon tanpa pemotongan menghasilkan nilai rata-rata precision 63.93%, recall 60.73%, dan akurasi 60.52%. Sedangkan pohon keputusan yang dipotong dengan menggunakan metode error-based pruning dengan menggunakan nilai confidence 0,4 menghasilkan precision 70.70%, recall 50.65%, dan akurasi 61.57%. Pohon yang dipotong dengan menggunakan nilai confidence 0,25 menghasilkan precision 73.77%, recall 48.84%, dan akurasi 62.44%. Dari hasil tersebut dapat disimpulkan bahwa penggunaan nilai confidence 0,25 meningkatkan akurasi lebih baik daripada nilai confidence 0,4 [10]. Berdasarkan latar belakang tersebut diharapkan hasil penelitian agar lebih mudah dipahami dan dapat dijadikan sebagai pembelajaran untuk kedepannya dalam menghasilkan para calon mahasiswa baru yang bermutu.

## 2. METODOLOGI PENELITIAN

### 2.1 Rancangan Penelitian

Penelitian ini dilakukan berdasarkan rumusan masalah yang telah dijabarkan pada bab sebelumnya yaitu untuk menentukan klasifikasi penerimaan calon mahasiswa baru. Penelitian ini dilakukan di STIKOM Tunas Bangsa Pematangsiantar, Sumatra Utara, dengan melakukan observasi dan wawancara secara langsung kepada pihak Penerimaan Mahasiswa Baru (PMB). Adapun rancangan penelitian yang digunakan dapat dilihat pada gambar dibawah.



**Gambar 1.** Kerangka Kerja Penelitian

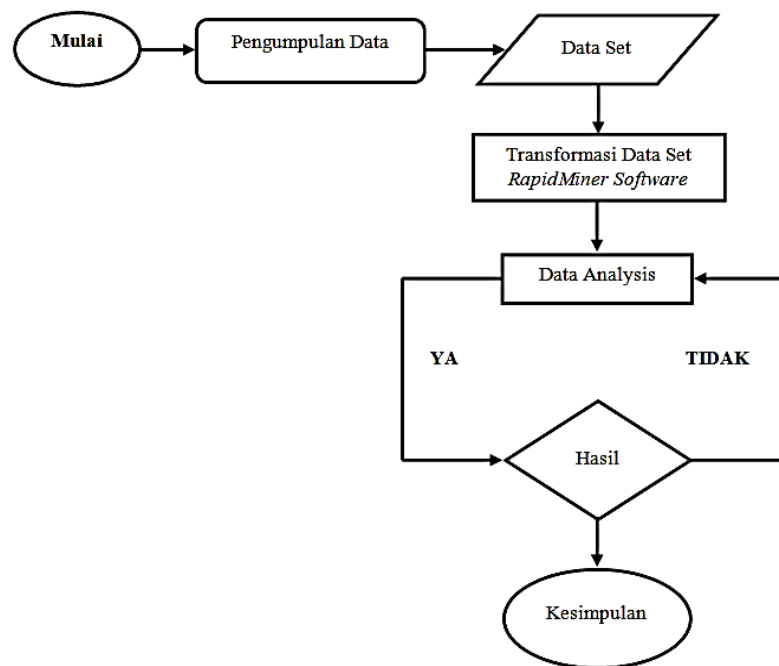
## 2.2 Pengumpulan Data

Pada klasifikasi penerimaan calon mahasiswa baru diproses dengan membuat pohon keputusan untuk menghasilkan suatu *output*. Di mana data tersebut mempunyai beberapa kriteria yang merupakan syarat dalam pengklasifikasian penerimaan calon mahasiswa baru. Data dikumpulkan dengan dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui secara jelas. Adapun data-data yang diperlukan dalam penelitian ini adalah data penerimaan calon Mahasiswa baru antara lain:

- Data penerimaan calon mahasiswa baru berdasarkan Nilai Ujian Tertulis Mahasiswa baru
- Data penerimaan calon mahasiswa baru berdasarkan kehadiran wawancara
- Data penerimaan calon mahasiswa baru berdasarkan nilai ujian praktek
- Data penerimaan calon mahasiswa baru berdasarkan nilai rata-rata UAN.

## 2.3 Diagram Aktifitas Kerja Penelitian

Berikut ini alur kerja yang akan dilakukan pada penelitian ini yang digambarkan dalam diagram aktivitas yang menjelaskan proses mengidentifikasi masalah dan tujuan penelitian yang dilakukan pada aplikasi RapidMiner seperti yang ditunjukkan pada gambar 2 berikut ini.



**Gambar 2.** Activity Diagram Penelitian

## 2.4 Pengolahan Data

Dari hasil proses pengumpulan data yang diperoleh dari panitia penerimaan mahasiswa baru, ditemukan komponen variabel yang selanjutnya akan digunakan sebagai data penelitian. Variabel tersebut adalah sebagai berikut.

- NIM
- Jenis Kelamin
- Status Mahasiswa
- Asal Sekolah
- Tahun Lulus Sekolah
- Program Studi
- Kelas
- Indeks Prestasi Kumulatif
- Nilai TPA
- Nilai Wawancara

Sebelum melakukan pengolahan data, terlebih dahulu dilakukan tahapan *pre-processing* data. *Pre-processing* data bertujuan untuk memperbaiki data apabila terdapat *missing values* atau kesalahan ketik. Pada penelitian, ini normalisasi data yang dilakukan adalah menghapus data yang tidak memiliki informasi lengkap dan tidak dibutuhkan dalam penelitian. Setelah data yang dianggap tidak dibutuhkan dalam penelitian ini dihilangkan, Langkah selanjutnya yang perlu dilakukan adalah melakukan transformasi data. Transformasi data yang dilakukan dalam penelitian ini adalah membentuk kelas setiap atribut yang tidak memiliki kelas. Atribut pada variabel data yang dilakukan transformasi data adalah sebagai berikut:

a) Asal Sekolah

Banyaknya atribut untuk variabel ASAL SEKOLAH dianggap perlu dilakukan transformasi. Hal ini dilakukan untuk mempersingkat pencabangan pada pohon keputusan yang dibentuk. Adapun hasil transformasi variabel asal sekolah dikelompokkan atas :

- 1) SMA, untuk mahasiswa yang berasal dari jenjang Sekolah Menengah Atas.
- 2) SMK, untuk mahasiswa yang berasal dari jenjang Sekolah Menengah Kejuruan.
- 3) MA, untuk mahasiswa yang berasal dari jenjang Madrasah Aliyah.

b) Tahun Lulus Sekolah

Variabel Tahun Lulus Sekolah diperoleh dari transformasi pengurangan tahun pendaftaran sebagai mahasiswa dikurang dengan tahun kelulusan sekolah.

c) Indeks Prestasi Kumulatif (IPK)

IPK merupakan nilai rata-rata seluruh mata kuliah yang telah ditempuh sejak semester pertama sampai akhir. Variabel ini memiliki nilai atribut *continuous*, artinya memiliki rentang nilai desimal dari 0 sampai dengan nilai 4. Dikarenakan rentang nilai yang cukup banyak, maka dilakukan klasifikasi terhadap nilai setiap atribut berdasarkan :

- 1) KURANG, untuk nilai IPK antara 0 sampai 2.00
- 2) BAIK, untuk rentang nilai IPK yang diperoleh antara 2.00 sampai dengan 3.00
- 3) SANGAT BAIK, untuk rentang nilai IPK yang diperoleh antara 3.00 sampai dengan 3.80
- 4) CUMLAUDE, untuk rentang nilai IPK 3.80 sampai 4.00

Sampel data setelah dilakukan *pre-processing* adalah sebagai berikut:

**Tabel 1.** Data setelah *Pre-Processing*

| NIM           | Jenkel | Asal Sekolah | Masuk Kuliah | Prodi | Kelas | IPK         |
|---------------|--------|--------------|--------------|-------|-------|-------------|
| 1602027       | Wanita | SMA          | 1 Tahun      | SI    | Pagi  | Sangat Baik |
| 1602165       | Pria   | SMK          | 1 Tahun      | SI    | Malam | Baik        |
| 1602187       | Pria   | SMK          | 2 Tahun      | SI    | Malam | Kurang      |
| 1602186       | Pria   | MA           | 2 Tahun      | SI    | Malam | Sangat Baik |
| 1602147       | Pria   | SMA          | 1 Tahun      | SI    | Malam | Sangat Baik |
| 1602304       | Pria   | MA           | 2-4 Tahun    | TI    | Siang | Sangat Baik |
| 1601003       | Pria   | MA           | 1 Tahun      | TI    | Pagi  | Sangat Baik |
| 1602323       | Wanita | SMK          | 1 Tahun      | SI    | Pagi  | Sangat Baik |
| 1602128       | Pria   | MA           | 1 Tahun      | SI    | Siang | Kurang      |
| 1602157       | Wanita | SMK          | 1 Tahun      | SI    | Malam | Baik        |
| 1602301       | Pria   | SMA          | 1 Tahun      | TI    | Siang | Baik        |
| 1602331       | Pria   | SMK          | 2 Tahun      | SI    | Pagi  | Baik        |
| 1602058       | Wanita | MA           | 0 Tahun      | SI    | Siang | Sangat Baik |
| 1602002       | Pria   | SMA          | 1 Tahun      | SI    | Pagi  | Baik        |
| 1602160       | Pria   | SMK          | 1 Tahun      | SI    | Malam | Sangat Baik |
| 1602046       | Wanita | SMK          | 0 Tahun      | SI    | Pagi  | Sangat Baik |
| 1601098       | Pria   | SMK          | 1 Tahun      | TI    | Siang | Baik        |
| 1602055       | Pria   | MA           | 0 Tahun      | SI    | Siang | Sangat Baik |
| 1602294       | Pria   | SMK          | 2 Tahun      | TI    | Siang | Sangat Baik |
| 1602205       | Wanita | SMA          | 1 Tahun      | SI    | Siang | Sangat Baik |
| <i>Dst...</i> |        |              |              |       |       |             |

Agar dapat diolah dalam *Software Rapidminer*, pertama kali data yang sudah di olah sebelumnya disimpan dalam format *microsoft excell* file berekstensi \*.xls.

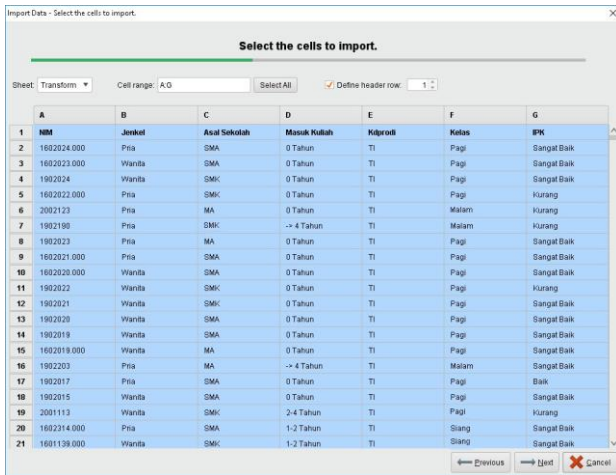
### 3. HASIL DAN PEMBAHASAN

Untuk memperoleh aturan pohon keputusan pada algoritma C.45, langkah-langkah yang dilakukan adalah:

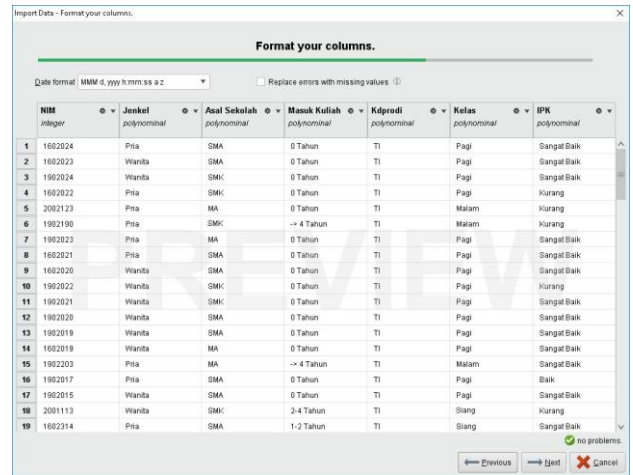
- a) Memilih atribut sebagai akar. Sebuah akar didapat dari nilai *gain* tertinggi dari atribut-atribut yang ada.
- b) Membuat cabang untuk masing-masing nilai,
- c) Membagi kasus dalam cabang, dan
- d) Mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Agar dapat diolah dalam *Software Rapidminer*, pertama kali data yang sudah di olah sebelumnya disimpan dalam format *microsoft excell* file berekstensi \*.xls.

- 1) Sesi Pemilihan Cell untuk di Import (Tahap 2) dan pemilihan variabel seperti yang ditunjukkan Gambar 3 dan 4 berikut.

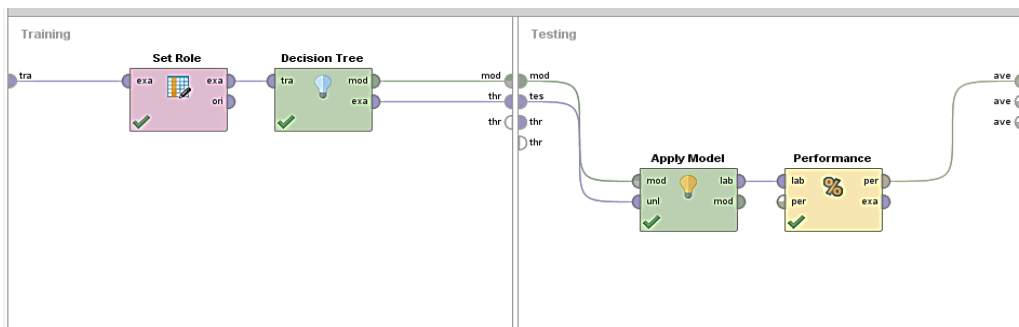


**Gambar 3.** Pemilihan Cell untuk di Import



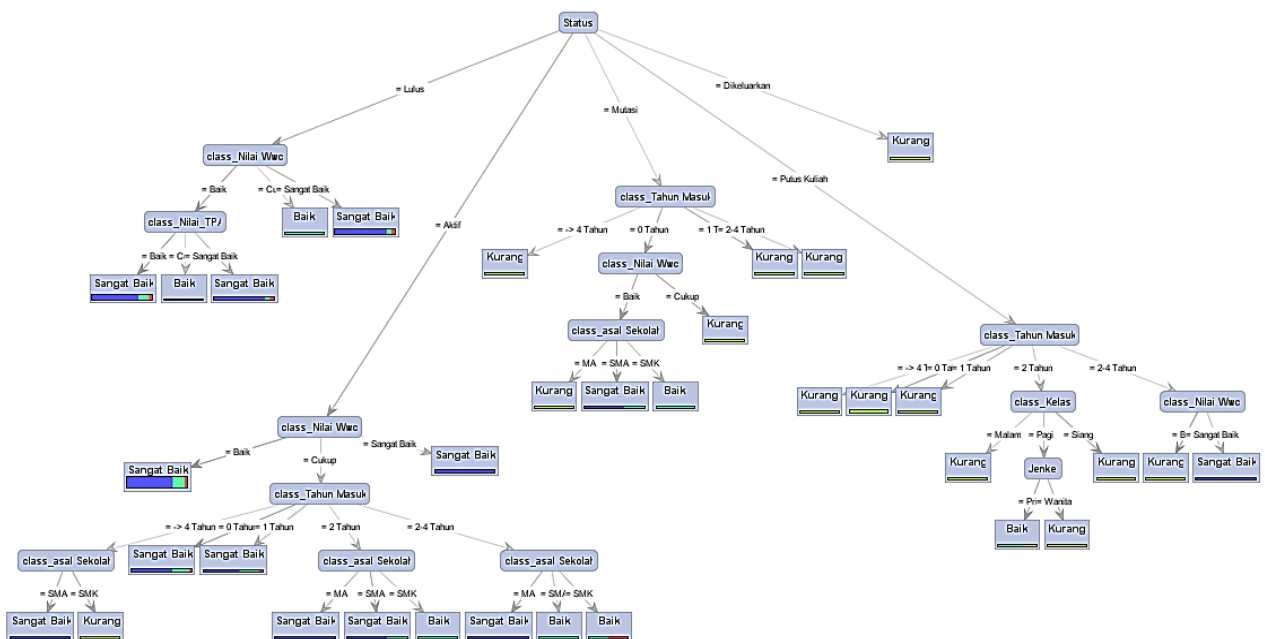
**Gambar 4.** Pemilihan Variabel sebagai Label

2) Tampil halaman operator validation yang terbagi atas area training dan testing yang ditunjukkan pada Gambar 5 berikut:



**Gambar 5.** Hasil keputusan pada views Result

Setelah seluruh langkah telah dikerjakan, *RapidMiner* akan melakukan proses dan menampilkan hasil keputusan pada views *Result*. Hasilnya berbentuk pohon keputusan seperti pada Gambar 6 berikut.



**Gambar 6.** Pohon keputusan

Dapat disimpulkan bahwa, kriteria pertama pada hasil prestasi mahasiswa adalah nilai wawancara, kemudian diikuti oleh hasil nilai pada saat ujian TPA. Sehingga, dapat diketahui bahwa terdapat hubungan yang linear antara nilai

Wawancara, Nilai TPA dan Prestasi terhadap kualitas mahasiswa. Semakin baik nilai yang diperoleh saat wawancara dan TPA, maka semakin baik prestasi yang diperoleh saat mengikuti perkuliahan. Dari grafik pohon keputusan yang diperoleh tersebut, dapat diketahui juga bahwa mahasiswa yang PUTUS KULIAH, MUTASI, dan mahasiswa yang DIKELUARKAN, cenderung memiliki nilai wawancara KURANG saat seleksi penerimaan mahasiswa.

- 1) Jika Nilai Wawancara = BAIK maka prestasi calon mahasiswa setelah mengikuti perkuliahan adalah SANGAT BAIK. Hasil yang diperoleh {Sangat BAIK=552, Baik=139, KURANG=0, CUMLAUDE=29}
- 2) Jika Nilai Wawancara = SANGAT BAIK maka prestasi calon mahasiswa setelah mengikuti perkuliahan SANGAT BAIK. Hasil yang diperoleh SANGAT BAIK=111, BAIK=0, Kurang=0, CUMLAUDE=0}
- 3) Jika Nilai Wawancara = CUKUP maka prestasi calon mahasiswa setelah mengikuti perkuliahan ditentukan oleh lamanya MASUK KULIAH.
  - a) Jika TAHUN MASUK kuliah = 0 Tahun, maka Prestasi calon mahasiswa setelah mengikuti perkuliahan SANGAT BAIK {SANGAT BAIK=70, BAIK=29, KURANG=0, CUMLAUDE=4}
  - b) Jika TAHUN MASUK kuliah = 1 Tahun, maka prestasi calon mahasiswa setelah mengikuti perkuliahan SANGAT BAIK {SANGAT BAIK=23, BAIK=12, KURANG=0, CUMLAUDE=3}
  - c) Jika TAHUN MASUK kuliah = 2 s.d 4 Tahun, maka prestasi calon mahasiswa akan dipengaruhi oleh asal sekolah.
- 4) Sementara itu, pada Rule mahasiswa yang DIKELUARKAN: diketahui bahwa Prestasi mahasiswa setelah mengikuti perkuliahan adalah KURANG {Sangat Baik=0, Baik=0, Kurang=48, Cumlaude=0}

Berikut adalah informasi sebaran berdasarkan IPK mahasiswa seperti yang tertera pada gambar berikut.

| Index | Nominal value | Absolute count | Fraction |
|-------|---------------|----------------|----------|
| 1     | Sangat Baik   | 1350           | 0.686    |
| 2     | Baik          | 296            | 0.150    |
| 3     | Kurang        | 241            | 0.122    |
| 4     | Cumlaude      | 82             | 0.042    |

**Gambar 7.** Distribusi Indeks Prestasi Mahasiswa

Mahasiswa yang memiliki IPK CUMLAUDE sebanyak 82 orang, SANGAT BAIK berjumlah 1350 orang, yang memiliki IPK BAIK sebanyak 296 orang, dan yang memiliki IPK KURANG baik sebanyak 296 orang.

classification\_error: 18.68% +/- 4.65% (mikro: 18.69%)

|                   | true Sangat Baik | true Baik | true Kurang | true Cumlaude | class precision |
|-------------------|------------------|-----------|-------------|---------------|-----------------|
| pred. Sangat Baik | 1347             | 271       | 4           | 81            | 79.10%          |
| pred. Baik        | 1                | 19        | 2           | 1             | 82.61%          |
| pred. Kurang      | 2                | 5         | 235         | 0             | 97.11%          |
| pred. Cumlaude    | 0                | 1         | 0           | 0             | 0.00%           |
| class recall      | 99.78%           | 6.42%     | 97.51%      | 0.00%         |                 |

**Gambar 8.** Nilai Akurasi Decision Tree

Dari hasil tersebut ditarik kesimpulan bahwa label mahasiswa yang memiliki IPK SANGAT BAIK mempunyai jumlah 1347 dengan *Class Precision* sebesar 79,10% dan prediksi label KURANG mempunyai nilai 2 dengan *Class Precision* sebesar 97,11%. Pengolahan data dengan metode *data mining* yang telah dilakukan dapat diketahui bahwa calon mahasiswa yang mendapatkan nilai wawancara SANGAT BAIK dan BAIK, berpengaruh terhadap prestasi saat mengikuti kuliah kedepannya. Mahasiswa yang memiliki nilai KURANG cenderung PUTUS KULIAH, MUTASI atau DIKELUARKAN.

## 4. KESIMPULAN

Dari grafik pohon keputusan yang diperoleh pada pengujian dengan *software RapidminerStudio*, dapat diketahui bahwa seorang calon mahasiswa akan aktif, lulus tepat waktu, berprestasi atau tidak setelah nantinya mengikuti perkuliahan, pertama kali dipengaruhi oleh faktor nilai wawancara. Dengan demikian dapat disimpulkan terdapat korelasi antara keaktifan calon mahasiswa, lulus tepat waktu, atau memiliki prestasi yang baik nantinya pada saat wawancara seleksi calon mahasiswa baru. Dari hasil penelitian, seorang calon mahasiswa yang memperoleh nilai wawancara SANGAT



BAIK pada saat seleksi penerimaan mahasiswa baru cenderung memiliki prestasi yang SANGAT BAIK setelah mengikuti perkuliahan. Sementara itu rata-rata calon mahasiswa yang mendapatkan hasil wawancara CUKUP saat seleksi penerimaan, akhirnya memiliki Indeks Prestasi Kumulatif BAIK setelah lulus kuliah. Berdasarkan pengolahan data menggunakan *software RapidMiner*, didapat nilai *accuracy* sebesar 81.32%, artinya bahwa rule yang dihasilkan tingkat kebenarannya sudah sangat baik. Dapat disimpulkan bahwa untuk melakukan klasifikasi pada penerimaan calon mahasiswa STIKOM Tunas Bangsa adalah nilai wawancara, kemudian diikuti oleh hasil nilai pada saat ujian TPA dan lamanya masuk kuliah. Terdapat korelasi antara nilai Wawancara, Nilai TPA, lamanya masuk kuliah dan Prestasi. Semakin baik nilai yang diperoleh saat wawancara, keaktifan dan prestasi yang diperoleh pada saat mengikuti perkuliahan kedepannya cenderung semakin baik. Dengan demikian, agar dapat meningkatkan kualitas mahasiswa dan lulusan STIKOM Tunas Bangsa, maka hal yang diperhatikan saat seleksi penerimaan mahasiswa baru adalah nilai wawancara yang diperoleh oleh calon mahasiswa. Dari hasil percobaan yang telah diperoleh juga dapat diketahui bahwa lamanya melanjutkan studi ke perguruan tinggi juga mempengaruhi prestasi calon mahasiswa setelah mengikuti perkuliahan. Jika pada saat mendaftarkan diri MASUK KULIAH diatas >4 tahun setelah lulus sekolah, prestasi yang diperoleh setelah mengikuti perkuliahan cenderung kurang baik dan Mutasi.

## REFERENCES

- [1] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [2] Taufiq and Y. Yudihartanti, "Penerapan Algoritma C4.5 Klasifikasi Tingkat Kelulusan Mahasiswa," *Semin. Nas. Ilmu Komput.*, vol. 2, pp. 153–162, 2019.
- [3] A. Rohman and A. Rufiyanto, "Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas," *Proceeding SINTAK 2019*, pp. 134–139, 2019.
- [4] A. H. Nasrullah, "Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 244–250, 2018, doi: 10.33096/ilkom.v10i2.300.244-250.
- [5] S. Takalapeta, "Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 3, pp. 34–38, 2018, doi: 10.37438/jimp.v3i3.186.
- [6] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *JOIN*, vol. 2, no. 1, pp. 36–41, 2017.
- [7] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [8] S. Sularno and P. Anggraini, "Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Keganasan Hama Pada Tanaman Padi (Studi Kasus : Dinas Pertanian Kabupaten Kerinci)," *J. Sains dan Inform.*, vol. 3, no. 2, p. 161, 2017, doi: 10.22216/jsi.v3i2.2779.
- [9] D. Ardiansyah, "Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, 2019, doi: 10.46846/jurnalinkofar.v1i2.29.
- [10] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, p. 1, 2018, doi: 10.23917/khif.v4i1.5975.